

The logarithmic least squares optimality of the geometric mean of weight vectors calculated from all spanning trees for (in)complete pairwise comparison matrices

Sándor Bozóki^{1,2}, Vitaliy Tsyganok^{3,4}

Abstract

Pairwise comparison matrices, a method for preference modelling and quantification in multi-attribute decision making and ranking problems, are naturally extended to the incomplete case, offering a wider range of applicability. The weighting problem is to find a weight vector that reflects the decision maker's preferences as well as possible. The logarithmic least squares problem has a unique and simply computable solution. The spanning tree approach does not assume any metric in advance, instead it goes through all minimal sufficient subsets (spanning trees) of the set of pairwise comparisons, then the weight vectors are aggregated. It is shown that the geometric mean of weight vectors, calculated from all spanning trees, is the optimal solution of the well known logarithmic least squares problem, not only for complete, as it was recently proved by Lundy, Siraj and Greco, but for incomplete pairwise comparison matrices as well.

1 Incomplete pairwise comparison matrices

Cardinal preferences of decision makers are often modelled and calculated by pairwise comparison matrices [16]. Questions 'How many times a criterion is more important than another one?' or 'How many times a given alternative is better than another one with respect to a fixed criterion?' are typical in multi-attribute decision problems. The numerical answers are collected into a pairwise comparison matrix $\mathbf{A} = [a_{ij}]_{i,j=1\dots n}$ having reciprocity, i.e., $a_{ij} = 1/a_{ji}$. A pairwise comparison matrix can be complete, as in a classical AHP model [16], or incomplete [11, 13]. In the paper incomplete means 'not necessarily complete', in other words, the number of missing elements is allowed to be zero.

Incomplete pairwise comparison matrices are applied not only in the same decision situations in which the complete matrices arise, but also larger decision and ranking problems. Bozóki, Csató and Temesi [1] propose a ranking method for top tennis players based on their pairwise results, where incompleteness occurs in a natural way. Csató [6] constructed a 149×149 incomplete pairwise comparison matrix to rank the teams of the 39th Chess Olympiad 2010.

¹Laboratory on Engineering and Management Intelligence, Research Group of Operations Research and Decision Systems, Institute for Computer Science and Control, Hungarian Academy of Sciences; Mail: 1518 Budapest, P.O. Box 63, Hungary. E-mail: bozoki.sandor@sztaki.mta.hu

²Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest, Hungary

³Laboratory for Decision Support Systems, Institute for Information Recording of National Academy of Sciences of Ukraine; Mail: 2, Shpak str., Kyiv, 03113, Ukraine. E-mail: tsyganok@ipri.kiev.ua

⁴Department of System Analysis, State University of Telecommunications, Ukraine

Example 1.1. Let \mathbf{A} be a 6×6 incomplete pairwise comparison matrix as follows:

$$\mathbf{A} = \begin{pmatrix} 1 & a_{12} & & a_{14} & a_{15} & a_{16} \\ a_{21} & 1 & a_{23} & & & \\ & a_{32} & 1 & a_{34} & & \\ a_{41} & & a_{43} & 1 & a_{45} & \\ a_{51} & & & a_{54} & 1 & \\ a_{61} & & & & & 1 \end{pmatrix}$$

2 The logarithmic least squares (LLS) problem

The basic problem of finding the best weight vectors usually includes an additional information on how closeness is defined or specified. The classical approaches apply metrics based on least squares [4], weighted least squares [4], logarithmic least squares [5, 12], just to name a few. A lot of further weighting methods are discussed by Golany and Kress [10] and by Choo and Wedley [3]. Even the well known eigenvector method [16] is proved to be a distance minimizing method [7, 8], although its metric seems to be rather artificial.

The Logarithmic Least Squares (LLS) problem [13] is defined as follows:

$$\min \sum_{\substack{i, j : \\ a_{ij} \text{ is known}}} \left[\log a_{ij} - \log \left(\frac{w_i}{w_j} \right) \right]^2 \quad (1)$$

$$w_i > 0, \quad i = 1, 2, \dots, n. \quad (2)$$

Originally, the LLS problem was defined for complete pairwise comparison matrices, i.e., the sum in the objective function is taken for all i, j . In this special case, the LLS optimal solution is unique and it can be explicitly computed by taking the geometric mean of rows' elements [5, 12].

The most common normalizations are $\sum_{i=1}^n w_i = 1$ and $\prod_{i=1}^n w_i = 1$. Normalization $w_1 = 1$, called *ideal-mode* in Lundy, Siraj and Greco [14]), can also be interpreted: the first object (criterion, alternative) is considered a reference point and all the others are expressed according to it.

Given an (in)complete pairwise comparison matrix \mathbf{A} of size $n \times n$, an undirected graph $G(V, E)$ is defined as follows: G has n nodes and the edge between nodes i and j is drawn if and only if the matrix element a_{ij} is known.

The incomplete LLS problem can be solved by the following theorem:

Theorem 2.1. (Bozóki, Fülöp, Rónyai [2]) *Let \mathbf{A} be an incomplete or complete pairwise comparison matrix such that its associated graph G is connected. Then the optimal solution $\mathbf{w} = \exp \mathbf{y}$ of the logarithmic least squares problem is the unique solution of the following system of linear equations:*

$$(\mathbf{L}\mathbf{y})_i = \sum_{k: e(i,k) \in E(G)} \log a_{ik} \quad \text{for all } i = 1, 2, \dots, n-1, n \quad (3)$$

$$y_1 = 0 \quad (4)$$

where \mathbf{L} denotes the Laplacian matrix of G (ℓ_{ii} is the degree of node i and $\ell_{ij} = -1$ if nodes i and j are adjacent).

\mathbf{L} has rank $n - 1$. Normalization (4), being equivalent to $w_1 = 1$, plays a technical role only. It can be replaced by, e.g., the commonly used $\prod_{i=1}^n w_i = 1$ ($\Leftrightarrow \sum_{i=1}^n y_i = 0$). The computational complexity of solving the system of n linear equations (3)-(4) is at most $O(n^3)$ [22, Chapter 1].

Example 2.1. Let incomplete pairwise comparison matrix \mathbf{A} be the same as in Example 1.1. Equations (3) for $i = 1, 2, \dots, 6$ form the following system of linear equations:

$$\begin{pmatrix} 4 & -1 & 0 & -1 & -1 & -1 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ -1 & 0 & -1 & 3 & -1 & 0 \\ -1 & 0 & 0 & -1 & 2 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} \log a_{12} + \log a_{14} + \log a_{15} + \log a_{16} \\ \log a_{21} + \log a_{23} \\ \log a_{32} + \log a_{34} \\ \log a_{41} + \log a_{43} + \log a_{45} \\ \log a_{51} + \log a_{54} \\ \log a_{61} \end{pmatrix}$$

3 Aggregations of weight vectors calculated from all spanning trees

The spanning tree approach by Tsyganok [19, 20] does not assume any distance function or measure of closeness. The basic idea is that the set of pairwise comparisons is considered as the union of minimal, connected subsets, or, in graph theoretical terms, spanning trees. Let S denote the number of all spanning trees of graph G . Every spanning tree determines a unique weight vector, fitting on the corresponding subset of matrix elements perfectly. Given a spanning tree, the calculation of its associated weight vector requires $O(n)$ steps.

The number of spanning trees can be very large. In the special case of complete pairwise comparison matrices, the number of all spanning trees is $S = n^{n-2}$ by Cayley's theorem. Another extremal case is when the graph of the incomplete pairwise comparison matrix is itself a tree ($S = 1$). The enumeration of all spanning trees requires $O(n + m + nS)$ steps, with the algorithm of Gabow and Myers [9], where m denotes the number of edges in G .

The computational complexity of calculating all weight vectors, associated to the spanning trees, is $\max\{O(nS), O(n + m + nS)\}$ steps, where S , the number of spanning trees, is between 1 and n^{n-2} .

The most natural candidates for the aggregation of weight vectors calculated from all spanning trees are the arithmetic [17, 18, 19, 20] and the geometric means [14, 21].

The following theorem connects two weighting methods.

Theorem 3.1. (Lundy, Siraj and Greco [14]) *The geometric mean of weight vectors calculated from all spanning trees is logarithmic least squares optimal in case of complete pairwise comparison matrices.*

The rest of the paper provides the generalization of this result, it is shown that it holds for incomplete matrices as well.

4 Main result: the geometric mean of weight vectors calculated from all spanning trees is logarithmic least squares optimal

Theorem 4.1. *Let \mathbf{A} be an incomplete or complete pairwise comparison matrix such that its associated graph is connected. Then the optimal solution of the logarithmic least squares problem*

is equal, up to a scalar multiplier, to the geometric mean of weight vectors calculated from all spanning trees.

Proof. Let G be the connected graph associated to the (in)complete pairwise comparison matrix \mathbf{A} and let $E(G)$ denote the set of edges. The edge between nodes i and j is denoted by $e(i, j)$. The Laplacian matrix of graph G is denoted by \mathbf{L} . Let $T^1, T^2, \dots, T^S, \dots, T^S$ denote the spanning trees of G , where S denotes the number of spanning trees. $E(T^s)$ denotes the set of edges in T^s . Hereafter, upper index s is also used for indexing a weight vector or a pairwise comparison matrix, associated to spanning tree T^s . Let $\mathbf{w}^s, s = 1, 2, \dots, S$, denote the weight vector calculated from spanning tree T^s . Weight vector \mathbf{w}^s is unique up to a scalar multiplication. For sake of simplicity we can assume that $w_1^s = 1$, but other ways of normalization, e.g., $\prod w_i = 1$ can also be chosen. Let $\mathbf{y}^s := \log \mathbf{w}^s, s = 1, 2, \dots, S$, where the logarithm is taken element-wise. Let \mathbf{w}^{LLS} denote the optimal solution to the incomplete Logarithmic Least Squares problem (normalized by $w_1^{LLS} = 1$) and $\mathbf{y}^{LLS} := \log \mathbf{w}^{LLS}$, then by Theorem 2.1,

$$(\mathbf{L}\mathbf{y}^{LLS})_i = \sum_{k: e(i,k) \in E(G)} b_{ik} \quad \text{for all } i = 1, 2, \dots, n,$$

where $b_{ik} = \log a_{ik}$ for all $(i, k) \in E(G)$.

In order to prove the theorem, it is sufficient to show that

$$\left(\mathbf{L} \frac{1}{S} \sum_{s=1}^S \mathbf{y}^s \right)_i = \sum_{k: e(i,k) \in E(G)} b_{ik} \quad \text{for all } i = 1, 2, \dots, n. \quad (5)$$

Consider an arbitrary spanning tree T^s . Then $\frac{w_i^s}{w_j^s} = a_{ij}$ for all $e(i, j) \in E(T^s)$. Introduce the incomplete pairwise comparison matrix \mathbf{A}^s by $a_{ij}^s := a_{ij}$ for all $e(i, j) \in E(T^s)$ and $a_{ij}^s := \frac{w_i^s}{w_j^s}$ for all $e(i, j) \in E(G) \setminus E(T^s)$. Again, $b_{ij}^s := \log a_{ij}^s (= y_i^s - y_j^s)$. Note that the Laplacian matrices of \mathbf{A} and \mathbf{A}^s are the same (\mathbf{L}). Since weight vector \mathbf{w}^s is generated by the matrix elements belonging to spanning tree T^s , it is the optimal solution of the LLS problem regarding \mathbf{A}^s , too. Equivalently, the following system of linear equations holds.

$$(\mathbf{L}\mathbf{y}^s)_i = \sum_{k: e(i,k) \in E(T^s)} b_{ik} + \sum_{k: e(i,k) \in E(G) \setminus E(T^s)} b_{ik}^s \quad \text{for all } i = 1, 2, \dots, n. \quad (6)$$

Lemma 4.1.

$$\sum_{s=1}^S \left(\sum_{k: e(i,k) \in E(T^s)} b_{ik} + \sum_{k: e(i,k) \in E(G) \setminus E(T^s)} b_{ik}^s \right) = S \sum_{k: e(i,k) \in E(G)} b_{ik}. \quad (7)$$

Proof. Let i be fixed arbitrarily and consider node i in all spanning trees. There is nothing to do with edges $e(i, k) \in E(T^s)$. Since T^s is a spanning tree, for every edge $e(i, k) \in E(G) \setminus E(T^s)$ there exists a unique path

$P = \{e(i, k_1), e(k_1, k_2), \dots, e(k_\ell, k)\} \subseteq E(T^s)$. $P \cup e(i, k)$ is a cycle and

$$b_{ik}^s = b_{ik_1} + b_{k_1 k_2} + \dots + b_{k_\ell k}. \quad (8)$$

Consider the following spanning tree: $T^{s'_{i,k,k_1}} := (T^s \setminus e(i, k_1)) \cup e(i, k)$ as in Figure 1.

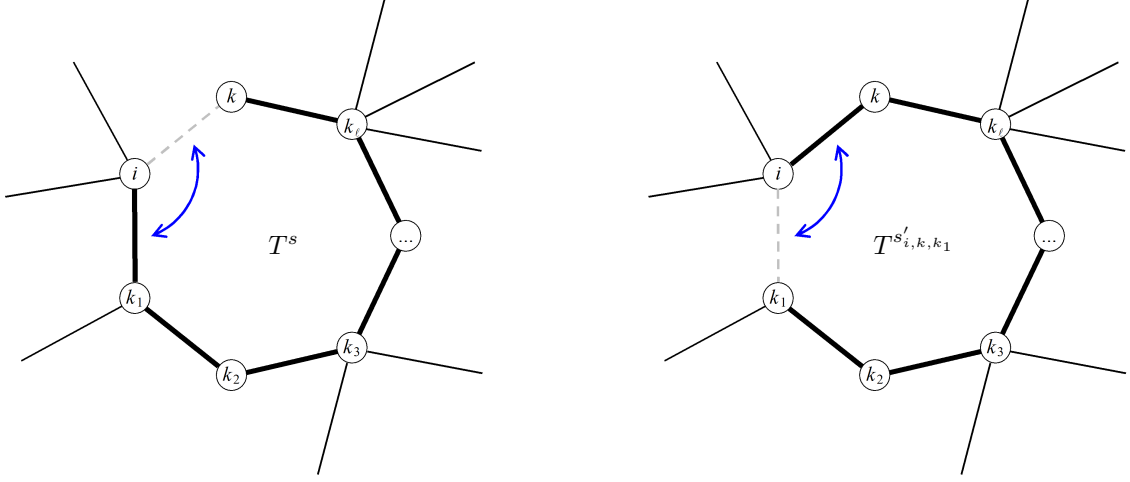


Figure 1. The replacement of edge $e(i, k_1)$ in spanning tree T^s by edge $e(i, k)$ results in spanning tree T^{s', k, k_1} .

Spanning trees T^s and T^{s', k, k_1} differ in one edge only and we can write again that

$$b_{ik_1}^{s', k, k_1} = b_{ik} + b_{kk_\ell} + \dots + b_{k_2 k_1}. \quad (9)$$

The sum of equations (8) and (9) results in

$$b_{ik}^s + b_{ik_1}^{s', k, k_1} = b_{ik} + b_{ik_1} \quad (10)$$

because all intermediate terms vanish due to the reciprocal property of pairwise comparison matrices. Now let us continue this process and go through all edges $e(i, k) \in E(G) \setminus E(T^s)$ for all k and s . The remarkable symmetry of the set of all spanning trees implies that every edge occurs in exactly one pair. Summing all these equations like (10), the statement of Lemma 4.1 follows. An illustrative example is given below in Example 4.1. \square

Finally, to complete the proof of Theorem 4.1, take the sum of equations (6) for all $s = 1, 2, \dots, S$ and apply Lemma 4.1 to conclude that $\mathbf{y}^{LLS} = \frac{1}{S} \sum_{s=1}^S \mathbf{y}^s$. \square

Remark. Complete pairwise comparison matrices ($S = n^{n-2}$) are included in Theorem 4.1 as a special case. The proof of Theorem 4.1 can also be considered as a second, and shorter proof of Theorem 3.1.

Example 4.1. (An illustration of Lemma 4.1) Let incomplete pairwise comparison matrix \mathbf{A} be the same as in Example 1.1. The associated graph G and its spanning trees T^1, T^2, \dots, T^{11} are drawn in Figure 2.

Let us focus on node $i = 1$. Edges departing from node 1 are missing 12 times (and they are not missing 32 times) in the whole set of spanning trees, hence can identify 6 pairs. They induce 6 pairs of equations, that are labelled in Figure 2. In tree T^1 ,

$$b_{12}^1 = b_{15} + b_{54} + b_{43} + b_{32} \quad (11)$$

Note that equation (11), as well as the forthcoming ones, are labelled on the corresponding edges in Figure 2. Now $s = 1, k = 2, k_1 = 5$ and $s'_{1,2,5} = 4$, because the replacement of edge $e(1, 5)$ in tree T^1 by edge $e(1, 2)$ results in tree T^4 . Here

$$b_{15}^4 = b_{12} + b_{23} + b_{34} + b_{45} \quad (12)$$

The sum of equations (11) and (12) confirms (10).

Let us continue by edge $e(1, 4)$ in tree T^1 .

$$b_{14}^1 = b_{15} + b_{54} \quad (13)$$

$$b_{15}^2 = b_{14} + b_{45} \quad (14)$$

The remaining four pairs of edges and their equations are listed below.

$$b_{12}^2 = b_{14} + b_{43} + b_{32} \quad (15)$$

$$b_{14}^4 = b_{12} + b_{23} + b_{34} \quad (16)$$

$$b_{12}^3 = b_{14} + b_{43} + b_{32} \quad (17)$$

$$b_{14}^7 = b_{12} + b_{23} + b_{34} \quad (18)$$

$$b_{14}^5 = b_{15} + b_{54} \quad (19)$$

$$b_{15}^8 = b_{14} + b_{45} \quad (20)$$

$$b_{14}^6 = b_{15} + b_{54} \quad (21)$$

$$b_{15}^9 = b_{14} + b_{45} \quad (22)$$

Lemma 4.1 is now confirmed for $i = 1$:

$$\sum_{s=1}^{11} \left(\sum_{k: e(1,k) \in E(T^s)} b_{1k} + \sum_{k: e(1,k) \in E(G) \setminus E(T^s)} b_{1k}^s \right) = 11 \sum_{k: e(1,k) \in E(G)} b_{1k} = 11(b_{12} + b_{14} + b_{15} + b_{16}).$$

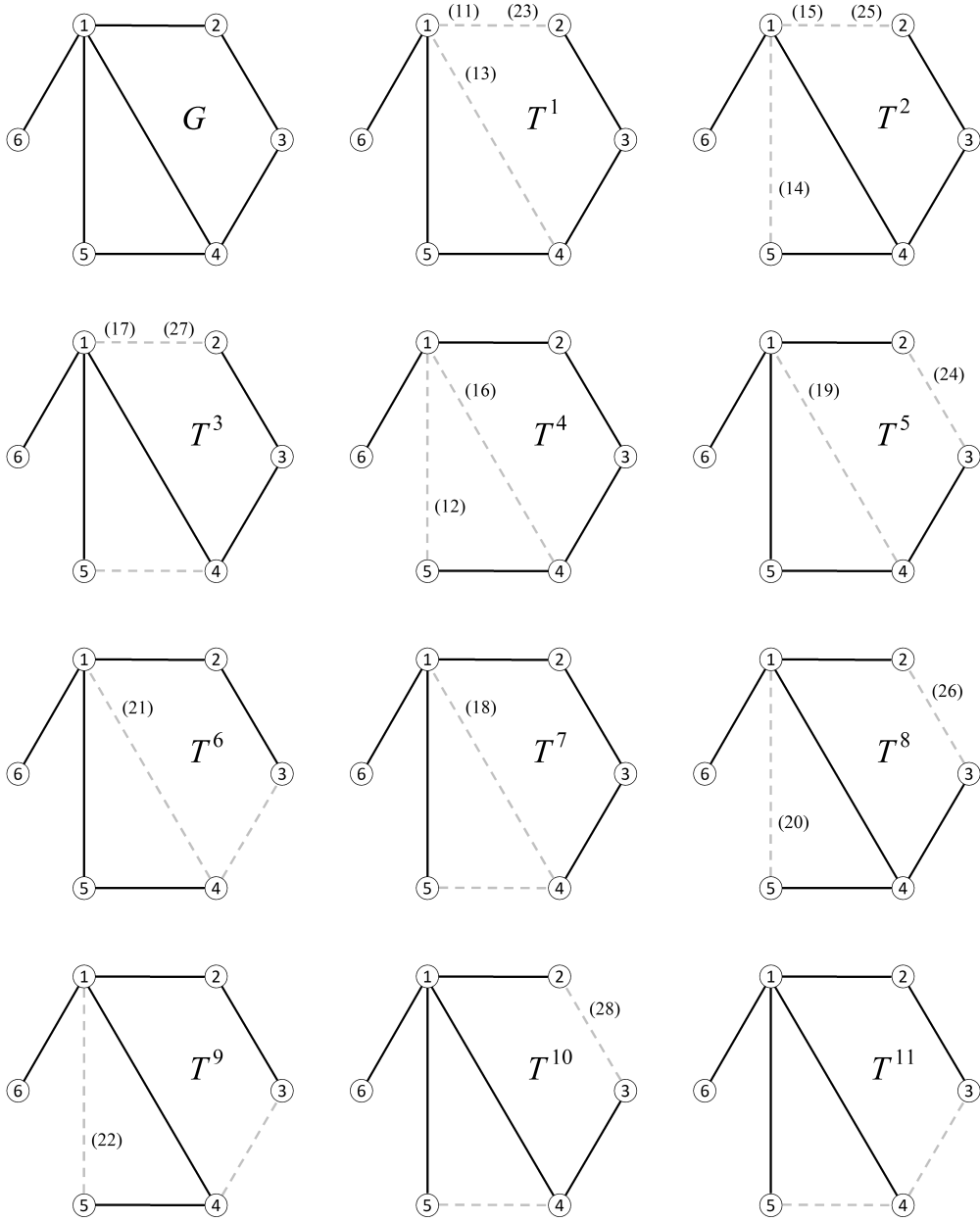


Figure 2. Graph G of Example 4.1 and its spanning trees T^1, T^2, \dots, T^{11}

Let us move to node 2, three pairs of equations are written:

$$b_{21}^1 = b_{23} + b_{34} + b_{45} + b_{51} \quad (23)$$

$$b_{23}^5 = b_{21} + b_{15} + b_{54} + b_{43} \quad (24)$$

$$b_{21}^2 = b_{23} + b_{34} + b_{41} \quad (25)$$

$$b_{23}^8 = b_{21} + b_{14} + b_{43} \quad (26)$$

$$b_{21}^3 = b_{23} + b_{34} + b_{41} \quad (27)$$

$$b_{23}^{10} = b_{21} + b_{14} + b_{43} \quad (28)$$

Lemma 4.1 is now confirmed for $i = 2$:

$$\sum_{s=1}^{11} \left(\sum_{k:e(2,k) \in E(T^s)} b_{2k} + \sum_{k:e(2,k) \in E(G) \setminus E(T^s)} b_{2k}^s \right) = 11 \sum_{k:e(2,k) \in E(G)} b_{2k} = 11(b_{21} + b_{23}).$$

The remaining nodes are left to the reader.

5 Conclusions

It has been shown in the paper that two weighting methods, based on rather different principles and approaches, are in fact equivalent not only for complete pairwise comparison matrices, as Lundy, Siraj and Greco [14] recently proved, but also for incomplete ones. The geometric mean of weight vectors calculated from all spanning trees has been proved to be logarithmic least squares optimal. The advantages rooted in the definition of the two methods, namely the clear interpretation of taking all spanning trees into account, and the optimality by a well understood objective function (LLS), have now been united.

There is a significant difference in computational complexity. The logarithmic least squares problem can be solved from a single system of linear equations (the coefficient matrix is the Laplacian), requiring at most $O(n^3)$ steps. The spanning tree approach requires $\max\{O(nS), O(n + m + nS)\}$ steps, where S , the number of spanning trees is between 1 and n^{n-2} . As soon as S exceeds $O(n^2)$, the logarithmic least squares problem is faster to solve.

An important consequence of the paper is that future analyses of weighting methods should not distinguish between the incomplete LLS and the geometric mean of weight vectors from all spanning trees.

Certain applications apply the spanning trees' enumeration, but not necessarily together with the aggregation by the geometric mean. The approach of spanning trees enumeration is used in determining the consistency to build the distribution of expert estimates based on the matrix [15]. Such problems offer further research possibilities.

The possible equivalence of the arithmetic (or other but not geometric) mean of weight vectors, calculated from all spanning trees, and other weighting methods, is still an open problem.

Acknowledgements

S. Bozóki acknowledges the support of the János Bolyai Research Fellowship no. BO/00154/16/3 and the Hungarian Scientific Research Fund (OTKA), grant no. K111797.

References

- [1] Bozóki, S., Csató, L., Temesi, J. (2016): An application of incomplete pairwise comparison matrices for ranking top tennis players, *European Journal of Operational Research*, 248(1) 211–218
- [2] Bozóki, S., Fülöp, J., Rónyai, L. (2010): On optimal completion of incomplete pairwise comparison matrices, *Mathematical and Computer Modelling*, 52(1-2):318–333
- [3] Choo, E.U., Wedley, W.C. (2004): A common framework for deriving preference values from pairwise comparison matrices, *Computers & Operations Research*, 31(6) 893–908
- [4] Chu, A.T.W., Kalaba, R.E., Spingarn, K. (1979): A comparison of two methods for determining the weights of belonging to fuzzy sets, *Journal of Optimization Theory and Applications*, 27(4) 531–538
- [5] Crawford, G., Williams, C. (1985): A note on the analysis of subjective judgment matrices, *Journal of Mathematical Psychology* 29(4) 387–405
- [6] Csató, L. (2013): Ranking by pairwise comparisons for Swiss-system tournaments, *Central European Journal of Operations Research* 21(4) 783–803
- [7] Fichtner, J. (1984). Some thoughts about the Mathematics of the Analytic Hierarchy Process. Report 8403, Universität der Bundeswehr München, Fakultät für Informatik, Institut für Angewandte Systemforschung und Operations Research, Werner-Heisenberg-Weg 39, D-8014 Neubiberg, F.R.G. 1984.
- [8] Fichtner, J. (1986). On deriving priority vectors from matrices of pairwise comparisons. *Socio-Economic Planning Sciences*, 20(6) 341–345
- [9] Gabow, H.N., Myers, E.W. (1978): Finding all spanning trees of directed and undirected graphs, *SIAM Journal on Computing*, 7(3) 280–287
- [10] Golany, B., Kress, M. (1993): A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices, *European Journal of Operational Research*, 69(2) 210–220
- [11] Harker, P.T. (1987): Incomplete pairwise comparisons in the analytic hierarchy process, *Mathematical Modelling* 9(11) 837–848
- [12] de Jong, P. (1984): A statistical approach to Saaty’s scaling methods for priorities, *Journal of Mathematical Psychology* 28(4) 467–478
- [13] Kwiesielewicz, M. (1996): The logarithmic least squares and the generalised pseudoinverse in estimating ratios, *European Journal of Operational Research* 93(3) 611–619
- [14] Lundy, M., Siraj, S., Greco, S. (2017): The mathematical equivalence of the “spanning tree” and row geometric mean preference vectors and its implications for preference analysis, *European Journal of Operational Research* 257(1) 197–208
- [15] Olenko, A., Tsyganok, V. (2016): Double Entropy Inter-Rater Agreement Indices, *Applied Psychological Measurement* 40(1) 37–55
- [16] Saaty, T.L. (1977): A scaling method for priorities in hierarchical structures, *Journal of Mathematical Psychology* 15(3) 234–281

- [17] Siraj, S., Mikhailov, L., Keane, J.A. (2012): Enumerating all spanning trees for pairwise comparisons, *Computers & Operations Research*, 39(2) 191–199
- [18] Siraj, S., Mikhailov, L., Keane, J.A. (2012): Corrigendum to “Enumerating all spanning trees for pairwise comparisons [Comput. Oper. Res. 39(2012) 191-199]”, *Computers & Operations Research*, 39(9) page 2265
- [19] Tsyganok, V. (2000): Combinatorial method of pairwise comparisons with feedback, *Data Recording, Storage & Processing* 2:92–102 (in Ukrainian).
- [20] Tsyganok, V. (2010): Investigation of the aggregation effectiveness of expert estimates obtained by the pairwise comparison method, *Mathematical and Computer Modelling*, 52(3-4) 538–54
- [21] Tsyganok, V.V., Kadenko, S.V., Andriichuk, O.V. (2015): Using different pair-wise comparison scales for developing industrial strategies, *International Journal of Management and Decision Making*, 14(3) 224–250
- [22] Watkins, D.S. (2002): *Fundamentals of Matrix Computations*, Second Edition. John Wiley & Sons, Inc., New York